



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Broad-spectrum Viral and Bacterial Pathogen Detection by Microarrays

K. McLoughlin, C. Jaing, S. Gardner, N. Be, J.
Thissen, T. Slezak

November 13, 2012

Microarrays: Current Technology, Innovations and Applications

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Book - Microarrays: Current Technology, Innovations and Applications

Publisher – Horizon Scientific Press

Chapter name: Broad spectrum viral and bacterial pathogen detection by microarrays

Running title: Pathogen detection microarray

Crystal Jaing^{*1}, Shea Gardner², Kevin McLoughlin², Nicholas A. Be¹, James B. Thissen¹, Tom Slezak²

¹Physical & Life Sciences Directorate, ²Computations Directorate, Lawrence Livermore National Laboratory, Livermore, CA



LLNL-BOOK-601460

*Author for correspondence

Crystal Jaing

jaing2@llnl.gov

Lawrence Livermore National Laboratory

P.O. Box 808

Mailstop L-452

Livermore, CA 94551

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Abstract

Broad spectrum viral and bacterial pathogen detection microarrays have proved to be a useful tool to analyze pathogenic organisms in clinical and environmental samples. The microbial detection arrays are lower cost and faster than next generation sequencing, and more comprehensive than singleplex and multiplex PCR. Pathogen detection arrays have been used for detection and genotyping of viral and bacterial pathogens for biodefense, public health, and food and drug safety. These arrays have also contributed to the discovery of novel viruses. Pathogen microarrays are cost-effective, rapid genomic technologies that are expected to gain a wider use both in research settings and in the regulatory and diagnostic environment.

Introduction

Rapid and accurate detection of viral and bacterial pathogens require technologies that are broad spectrum and cost-effective (Figure 1). Traditional detection technologies such as Polymerase chain reaction (PCR) detect small, prioritized sets of high-risk biological pathogens. Multiplex PCR analysis is relatively inexpensive and fast. It also has a high sensitivity for known organisms, but currently it can process no more than about 50 DNA signatures at one time, limited chiefly by unwanted combinatoric interactions between the primer sets. The likelihood of discovering unknown agents or new species is low with PCR analysis due to its high specificity, and viruses lack highly conserved regions for amplification by degenerate primers. At the other end of the cost spectrum, high-throughput sequencing provides the most comprehensive information about biological pathogens, both known and unknown. The process, however, is costly, can take several days to produce results, and requires dedicated bioinformatics capabilities to process and interpret large amounts of sequence data. Microarrays can identify a broad range of organisms, including pathogens on a priority screening list, sequenced bacteria or viruses that might not be anticipated, or even emerging pathogens containing DNA sequences similar to those previously identified in other pathogens. Processing samples on an array is currently less expensive and faster than sequencing, and more inclusive than multiplex PCR analysis.

Several different pathogen detection microarrays have been developed to analyze viral and bacterial agents from clinical and environmental samples (9, 16, 25, 28). In this chapter, we will review the methods to run the pathogen arrays, different pathogen detection array platforms, and discuss the advantages and challenges of microarray assays.

Description of Methods

A microarray is a miniaturized device containing short (25- to 70-mer) single-stranded DNA oligonucleotide probes (or “oligos”) attached to a solid substrate. While microarrays can be designed to perform many different tasks (e.g., resequencing, gene expression, etc.), we are focused here on hybridization arrays designed to detect the presence of nucleic acids from viral or other microbial organisms. The array probes are designed to have sequences complementary to segments of one or more target organism genomes. Oligos may be spotted onto the array by mechanical deposition (21), sprayed on with a modified inkjet printer head (10), or synthesized *in situ* through a series of photocatalyzed reactions (18). Probes are placed on the array in a rectangular grid of “features”, each containing many copies of the same oligo. The density of features on the array varies between platforms, from 20,000 spots per slide for a typical spotted array, to several million for platforms such as Roche NimbleGen, Agilent and Affymetrix that use *in situ* synthesized oligos (10, 22). Arrays may be subdivided with a gasket into subarrays, allowing multiple samples to be tested on one slide. Replicate features, scattered randomly across the array, may be used to allow correction for scratches and spatial effects.

On some arrays, negative control probes with random sequences or mismatches to targets are included, to provide a threshold level for background noise correction.

To perform a microarray experiment, first the nucleic acids including DNA and RNA need to be extracted from a sample. The sample could be environmental, clinical, tissue, blood or cell culture. There are different extraction kits commercially available to perform DNA and RNA extraction. After extraction, the nucleic acid samples are quantified using a spectrophotometer such as the NanoDrop spectrophotometer, a chip-based capillary electrophoresis instrument such as the Agilent Bioanalyzer, or fluorescent based quantification methods using RiboGreen or PicoGreen dyes. RNA samples are converted into cDNA by reverse transcription and then double-stranded cDNAs are generated using Klenow DNA polymerase. Very often, an environmental or a clinical sample will need to be amplified before microarray hybridization. Phi29 polymerase based whole genome amplification or random PCR- based amplification have been used successfully to detect viral and bacterial agents from clinical and environmental samples on microarrays (8, 11, 24).

The DNA and cDNA samples are labeled with fluorescent dyes before hybridization on a microarray to allow visualization of bound probes. There are a number of fluorescent dyes and kits commercially available for labeling. The labeling kit from Roche NimbleGen uses Cy3-labeled random nonamer primers and Klenow DNA polymerase at 37°C for 2 hr. The labeled DNA is then precipitated in isopropanol, and the pellet is washed, dried, reconstituted and quantified. Alternatively, a column based reagent can be used to clean up the labeled DNA. To load a NimbleGen array, 2-4 µg of labeled DNA is mixed with NimbleGen hybridization buffers and components and the arrays are allowed to hybridize on a hybridization station at 42°C for 16 hr. Arrays are then washed with wash buffers and scanned using a fluorescent scanner. Spotted microarray and Agilent arrays are processed similarly.

Overview of different pathogen detection microarrays

ViroChip

The first microarray designed for detection of a wide range of pathogens was the ViroChip (25). The initial version of the ViroChip contained 1,600 probes derived from the 140 complete viral genomes available in GenBank when the array was designed. Later versions of the array were developed to cover a wider range of viruses as additional genomes were published (1, 26, 29). The current version of the ViroChip, version 5, contains 60,000 probes representing complete and partial sequences of all viral genomes in GenBank as of December of 2010. The ViroChip was initially fabricated by mechanically spotting synthesized oligonucleotides on a glass slide, and has since been ported onto an Agilent inkjet platform. The oligos are 70mers, usually selected to match conserved sequences common to a taxonomic family.. For some families, oligos were also selected at the genus and species level to ensure uniform coverage across the viral taxonomy. Because the probes were designed against conserved sequences, the ViroChip could be used to identify novel viruses within the same family as a known, sequenced virus. This capability was used to characterize the virus

responsible for the 2003 SARS outbreak as a novel coronavirus (13), and to discover a number of novel viruses including a previously undescribed rhinovirus clade (12), human cardioviruses (4), and a novel adenovirus linked to a fulminant pneumonia outbreak in a New World monkey colony, with cross-species transmission to a human researcher (2). The ViroChip was also shown to be superior to direct fluorescent antibody testing and comparable to PCR in the diagnosis of respiratory infections in children (6), and due to its breadth of detection, was able to diagnose human parainfluenza type 4 and metapneumovirus infections in critically ill patients which had eluded conventional diagnostic testing in the hospital setting (3, 5).

The E-Predict algorithm was developed for analyzing ViroChip arrays (23). E-Predict computes a “theoretical hybridization energy profile” for each complete viral genome, by using BLAST to align probes to the genome sequence, and then computing a predicted hybridization free energy for each probe having a significant alignment. To be useful for analyzing clinical and environmental samples, a detection algorithm must be able to identify multiple organisms within a sample. This problem is addressed with an iterative version of E-Predict. After identifying the most likely target, E-Predict sets the intensities of the probes matching that target to zero, renormalizes the intensity vector, recomputes the similarity scores, and repeats the identification process. E-Predict has been successfully applied to identify or characterize viruses in thousands of ViroChip experiments. Another strategy that has been applied to detect viral signatures in ViroChip microarrays is assessing the significance of individual probes by Z-score analysis (3, 5). ViroChip intensity values are transformed into Z-scores by comparison with a “background” set of microarrays, and the most statistically significant viral probes are ranked in order to identify viruses which are most likely to be present in the sample.

GreeneChip

The “GreeneChip” arrays represent a broader-spectrum approach to pathogen detection (15, 19). These are high-density oligonucleotide arrays, fabricated using the Agilent inkjet system. GreeneChipVr version 1.0 contains 9,477 probes for viruses infecting vertebrates. GreeneChipPm v1.0 is a panmicrobial array design, containing all of the GreeneChipVr probes, together with probes for several thousand pathogenic bacteria, fungi and protozoa, comprising a total of 29,495 60-mer oligos. Viral probes were designed to target a minimum of 3 genomic regions for each family or genus of virus. Typically, one highly conserved region was chosen, along with 2 or more variable regions. Probe sequences were selected so that every vertebrate virus in the ICTV database (International Committee on Taxonomy of Viruses) or in GenBank was represented by at least one probe, with 5 or fewer mismatches. Bacterial, fungal and protozoan probes were selected by a similar strategy, except that the target sequences were only chosen from the 16S ribosomal RNA (rRNA) genes of bacteria and 18S rRNAs of fungi and protozoa. When they were tested with virus-infected cell cultures and clinical samples from virally infected patients, the GreeneChip arrays correctly identified the virus at the species level. Performance with bacterial samples was poorer, due to the choice of 16S rRNA as the target gene; probes for these targets tended to cross-hybridize across taxa, so that some bacteria could only be identified at family or

class resolution. The sensitivity of these arrays was comparable to that of the ViroChip series, due to the use of long (60-mer) oligos.

Many current algorithms for detection array analysis follow purely empirical approaches, without trying to model the physical processes underlying hybridization, washing and scanning. The algorithm developed for GreeneChip analysis, “log-transformed analysis of microarrays using p-values” (GreeneLAMP) (4), is one such approach to the species identification problem. The GreeneLAMP algorithm makes several key assumptions about array experiments: Probe intensities are log-normally distributed; probe intensities represent independent measurements of target genome concentrations; and the number of probes for any species having positive signals is limited, on the order of 100 or fewer.

When pairs of probe sequences are 95 percent or more identical, the independence assumption is clearly violated. In this case, the algorithm clusters probes into equivalence groups and pools their signals in an unspecified manner. Probes are associated with target taxa using BLAST; the score threshold for association is not specified, and there appears to be no attempt to differentiate between probes with strong and weak similarity. A potential limitation of GreeneLAMP is how it detects multiple organisms in complex mixtures, such as those found in clinical and environmental samples. Since the output of the algorithm is a single ranked list of taxa, there is no means to identify a combination of taxa that if present would best explain the observed intensity data.

Phylochip

Phylogenetic microarrays rely on probes which are complementary to sequences from bacterial small subunit rRNA, which are highly conserved. Such an approach is well suited to analyses which examine relative compositions within distinct microbial communities. The level of taxonomic specificity provided by these arrays varies dependent on probe selection, but depth of identification is inherently limited by use of these highly conserved sequences. Among the most commonly applied phylogenetic microarrays is the Phylochip (7).

The Phylochip is another prokaryotic detection array that exclusively probes 16S rRNA bacterial sequences using Affymetrix arrays with fixed-length 25-mer probes (7). It does not detect viruses, fungi, or protozoa, and is geared for detection and abundance studies of bacteria, allowing characterization at the genus or class level due to the poor resolution of 16S rRNA sequences at the species and strain level. It has been used extensively in environmental microbiology studies, although the low specificity precludes its application in disease surveillance, so we will not discuss it further here.

Higher specificity can be engendered by tailoring microarray design to a subset of pathogens representing a particular panel of interest. Microarrays may, for instance, be custom designed to identify pathogenic microorganisms prevalent in a certain physiological site within the host. A number of platforms have been developed toward this end, including the human intestinal tract chip (HITChip), which examines relative communities within human gastrointestinal flora (20). The HITChip was designed by extracting 16,000

rRNA sequences from intestinal bacteria and grouping these into 1,140 phylotypes, which were represented by 4,809 probes on the Agilent platform. Several other arrays with physiological site specificity have also been designed. A more complete outline of available phylogenetic microarrays, along with the breadth of their potential use, was effectively presented in a previous review, thus all possible applications of tailored arrays will not be outlined here (17).

Lawrence Livermore Microbial Detection Array (LLMDA)

The most comprehensive pathogen detection arrays reported to date were designed by a team at Lawrence Livermore National Laboratory (9, 11). Initial versions of the Lawrence Livermore Microbial Detection Array (LLMDA) contained 388,000 target probes for all bacteria and viruses (pathogenic and otherwise) for which full genome sequences were available from around 3200 species. More recent versions also include probes for pathogenic fungi, protozoa, and archaea as well as an updated set of viruses and bacteria, covering over 5,900 species with an average of over 100 probes per sequence element (complete genome, chromosome, viral segment, or plasmid). Probe lengths on a single array vary between 50 and 65 nucleotides, and are adjusted so that all probes on the array have roughly equivalent affinities for their complementary target DNA molecules. As on the GreeneChip, probes are selected from target genomes using two strategies. “Discovery” probes match genome regions that are unique to a taxonomic family or subfamily, but are shared by the species within that family. By targeting sequences that evolve more slowly within families, the discovery probes are optimized for detection of novel species within a known family. “Census” probes target highly variable regions that are unique to an individual species or strain. They are optimized for forensic use, to identify the specific strain of organism in a sample as precisely as possible.

The LLMDA designs are primarily deployed on the NimbleGen platform, and have also been tested with Agilent inkjet technology. NimbleGen arrays are fabricated by photocatalytic synthesis, using a digital micromirror device to control the addition of nucleotides to each oligo. This technology supports higher probe densities than either spotted oligo arrays or the Agilent inkjet platform, with up to 2.1 million features per array. Thus, the LLMDA is able to represent each sequenced microbial genome by a large number of diverse probes, varying from 10 to over 50 depending on the array format and the range of microbes targeted. A probe is counted toward the minimum representation of an organism if it has an alignment to the target genome with at least 85 percent identity over the length of the probe. Bacterial probes are selected from full genome sequences, rather than from 16S rRNA genes as in the GreeneChip, enabling strain-level discrimination for most bacteria.

The LLMDA was shown to have high sensitivity and specificity when tested against a variety of previously characterized viral and bacterial cultures, and was also used successfully to detect adventitious porcine circovirus DNA in a pediatric rotavirus vaccine (24). Recently, LLNL has established a diagnostic platform for unbiased random amplification and microarray identification of viral pathogens in clinical samples (8). This work showed that Phi29 polymerase-amplification of a diverse set of clinical samples generates

enough viral material for successful identification by the LLMDA, demonstrating the potential of the microarray technique for broad spectrum pathogen detection in human samples. It also showed that LLMDA detects both DNA and RNA viruses as well as bacteria and plasmids present in the same sample, and in some cases can differentiate different subtypes.

LLMDA processing begins by purifying DNA or RNA from a clinical, environmental or cell culture sample. Commercially available nucleic acid extraction kits such as the kits from Qiagen have been primarily used for sample preparation before LLMDA analysis. Extensive viral enrichment such as ultracentrifugation has not been routinely used for preparation of samples. For viral detection studies, filtration using a 0.22 μm filter to remove bacterial or mammalian cells, and DNase treatment have been used before LLMDA analysis (8), but this process has not been routinely conducted for unknown samples when both bacteria and viruses could be present. Additional mechanical and enzymatic protocols have also been applied to increase extraction yields for gram-positive bacteria.

Whole genome based amplification (WGA) using the Phi29 DNA polymerase or PCR with random primers have been used to amplify sample prior to LLMDA analysis (8, 24). Because the probes were designed from full genomic sequence, rather than 16S rRNA sequence, whole genome amplification is required for sensitive and specific detection of viral and bacterial agents. WGA does have the potential to alter relative representation of microbial communities in the resultant amplicons due to amplification bias. The LLMDA array, however, is generally applied in a detection capacity, and not as a quantitative population survey. Additionally, the effects of WGA have been demonstrated to be least prevalent using the Phi29-based REPLI-g system (27), the platform applied in conjunction with the LLMDA.

To find a combination of organisms whose presence in the sample best explains the observed data, LLNL developed the composite likelihood maximization algorithm (CLiMax), as described in (9, 14). The likelihood maximization algorithm is an iterative process, which scans repeatedly through a database of all sequenced microbial genomes. The first iteration looks for the target genome that explains the largest portion of the observed detected probe signals, while minimizing the number of probes expected to bind to target that did not have detectable signals. In each subsequent iteration, the algorithm chooses the target that explains the largest part of the signal not already explained by the first target, while again minimizing the number of expected negative probes. The process continues until a maximal portion of the observed probe signals are explained, or for a specified maximum number of iterations. An example of data from a sample analyzed by the LLMDA is shown in Figure 2. This is from a human clinical sample where

The CLiMax algorithm is implemented as part of an automated software pipeline, which incorporates web-based data entry, sample tracking and analysis. The microarray results are shown in a graphical format showing a list of predicted targets organized by viral or bacterial family, together with the log-likelihood scores for each target sequence.

Discussion

Broad-spectrum microarrays can be used for any applications where the rapid determination of known major bacterial and viral components is desired, and for detecting signals from novel near neighbors with homology to sequenced microbes. A comparison of resolution, computational time, throughput and cost is shown in Table 1. The significant advantages of the pathogen detection microarrays are broad coverage of bacteria and viruses, as well as eukaryotic pathogens. Other advantages of detection microarrays are lower cost relative to massively parallel sequencing, particularly if the bioinformatics analyses costs of sequencing versus arrays are tallied. For reagents alone, a virus microarray costs approximately \$50-\$500/sample depending on the format of the array, which is more expensive than multiplex PCR, but much less expensive than reagent costs of sequencing. Arrays are faster than most DNA sequencing platforms. Samples can be processed in under 24 hours from sample-in to results for arrays, while sequencing can involve lengthy library preparation up front and extensive computing on a large cluster to determine the organisms present in millions of short reads. When sequencing samples from a host, the majority of reads may be from the host, diluting signal and requiring deep coverage to detect microbes, and in these cases microarrays may be especially cost effective. Microarrays can be used to interrogate many sample types. Depending on the vendor, microarrays can process up to 12 samples at a time, with correspondingly decreased probe density and resolution. Microarrays provide a way to distribute a standardized quality across different labs for reliable detection of a broad range of microorganisms.

The one challenge of pathogen microarrays is determining the relative rate of false positives from microarray data. Certain probe sequences are “sticky” to various targets, but based on initial testing, these can be filtered out during data analysis. As with any technology based on nucleic acid detection, the capabilities of microarrays are limited by the genome sequence information available at the time of design. Many bacterial and viral species and strains are known but have not yet been sequenced. Microarrays cannot identify bacteria or viruses from families that have absolutely no genomes sequenced, nor can probe design strategies avoid selecting probes that may cross-hybridize to unsequenced organisms. All microarrays require regular re-design to take advantage of recent genomic sequence data. For truly novel organisms, sequencing is needed for detection and characterization.

Microarray analyses of unknown clinical and environmental samples often require random amplification. Potential amplification bias might prevent some array oligos from finding their target, or artificially increase hybridization signal from other oligos. Random amplification used with whole-genome arrays may yield lower sensitivity than arrays targeting a limited number of specific regions, for which specific PCR can be used. On the other hand, while specific PCR increases sensitivity for a limited number of targets, it also greatly limits the breadth of targets that can be detected.

Because the pathogen microarrays contain hundreds of thousands of probes, developing relevant standardized biological samples to validate specificity and sensitivity can be challenging. This relatively new technology is not yet broadly available for easy use and is not generally accepted as a diagnostic technology.

While pathogen detection arrays have been used primarily in a research context, several groups are aggressively working to develop arrays for clinical diagnostics, food safety testing, environmental monitoring and biodefense. It is noted that United States government agencies have yet to define formal validation procedures for high information content assays such as microarrays and genomic sequencing. This regulatory barrier may delay the widespread use of high-throughput techniques to identify pathogens in human clinical samples, food, vaccines and other products. However, food and pharmaceutical companies and diagnostic laboratories in the US and Europe have shown strong interest in applying these technologies. It is feasible that the attendant regulatory hurdles will eventually be overcome.

Conclusions

Broad-spectrum microarrays present a deterministic and demonstrated approach to identify known pathogens in environmental and clinical samples at detection levels that are relevant for health, safety and regulatory purposes. Compared to multiplex PCR assays and high-throughput sequencing, microarrays occupy an intermediate range of cost, processing time, and sensitivity for detection of microbial DNA in complex samples. Microarrays require random amplification of samples using random primers, instead of genome-specific amplification in a typical PCR assay. However, microarrays can query hundreds of thousands to several million regions of DNA in parallel (depending on platform and format), compared to at most a few tens of regions in the largest multiplexed PCR assays. However, PCR will continue to be, for the foreseeable future, the most sensitive discriminator of whether a known pathogen is present. Similarly, sequencing will remain the best way to determine the exact nucleotide composition for novel, mutated, or engineered pathogens. Nevertheless, we expect that microarrays will evolve during this decade to become a cost-effective means for health, drug, and food safety.

References

1. **Chen, E., S. Miller, J. DeRisi, and C. Chiu.** 2012. Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens. *PLoS Negl Trop Dis* **6**:e1485.
2. **Chen, E., S. Yagi, K. Kelly, S. Mendoza, R. Tarara, D. Canfield, N. Maninger, A. Rosenthal, A. Spinner, K. Bales, D. Schnurr, N. Lerche, and C. Chiu.** 2011. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS Pathog* **7**:e1002155.
3. **Chiu, C., A. Alizadeh, S. Rouskin, J. Merker, E. Yeh, S. Yagi, D. Schnurr, B. Patterson, D. Ganem, and J. DeRisi.** 2007. Diagnosis of a critical respiratory illness caused by human metapneumovirus by use of a pan-virus microarray. *J Clin Microbiol* **45**:2340-3.
4. **Chiu, C., A. Greninger, K. Kanada, T. Kwok, K. Fischer, C. Runckel, J. Louie, C. Glaser, S. Yagi, D. Schnurr, T. Haggerty, J. Parsonnet, D. Ganem, and J. DeRisi.** 2008. Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. *Proc Natl Acad Sci U S A* **105**:14124-9.
5. **Chiu, C., S. Rouskin, A. Koshy, A. Urisman, r. K. Fische, S. Yagi, D. Schnurr, P. Eckburg, L. Tompkins, B. Blackburn, J. Merker, B. Patterson, D. Ganem, and J. DeRisi.** 2006. Microarray detection of human parainfluenzavirus 4 infection associated with respiratory failure in an immunocompetent adult. *Clin Infect Dis* **43**:e71-6.
6. **Chiu, C., A. Urisman, T. Greenhow, S. Rouskin, S. Yagi, D. Schnurr, C. Wright, W. Drew, D. Wang, P. Weintrub, J. Derisi, and D. Ganem.** 2008. Utility of DNA microarrays for detection of viruses in acute respiratory tract infections in children. *J Pediatr* **153**:76-83.
7. **DeSantis, T., C. Stone, S. Murray, J. Moberg, and G. Andersen.** 2005. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiol Lett* **245**:271-8.
8. **Erlandsson, L., M. W. Rosenstierne, K. McLoughlin, C. Jaing, and A. Fomsgaard.** 2011. The Microbial Detection Array combined with random Phi29-amplification used as a diagnostic tool for virus detection in clinical samples. *PLoS ONE* **6**(8):e22631. doi:10.1371/journal.pone.0022631.
9. **Gardner, S., C. Jaing, K. McLoughlin, and T. Slezak.** 2010. A Microbial Detection Array (MDA) for viral and bacterial detection. *BMC Genomics* **11**:668 doi:10.1186/1471-2164-11-668.
10. **Hughes, T. R., M. Mao, A. R. Jones, J. Burchard, M. J. Marton, K. W. Shannon, S. M. Lefkowitz, M. Ziman, J. M. Schelter, M. R. Meyer, S. Kobayashi, and D. H. Davis C, He YD, Stephanians SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol.* **19**:342-7.
11. **Jaing, C., S. N. Gardner, K. McLoughlin, N. Mulakken, M. Alegria-Hartman, P. Banda, P. Williams, P. Gu, M. Wagner, C. Manohar, and T. R. Slezak.** 2008. A functional gene array for detection of bacterial virulence elements. *PLoS ONE* **3**(5):e2163. doi:10.1371/journal.pone.0002163.
12. **Kistler, A., P. Avila, S. Rouskin, D. Wang, T. Ward, S. Yagi, D. Schnurr, D. Ganem, J. DeRisi, and H. Boushey.** 2007. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* **196**:817-25.
13. **Ksiazek, T., D. Erdman, and C. Goldsmith.** 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**:1953-66.
14. **McLoughlin, K.** 2011. Microarrays for pathogen detection and analysis. *Briefings in Functional Genomics and Proteomics*:doi:10.1093/bfpg/elr027.
15. **Palacios, G., P. Quan, O. J. Jabado, S. Conlan, D. L. Hirschberg, Y. Liu, Zhai, J., N. Renwick, J. Hui, H. Hegyi, A. Grolla, J. E. Strong, J. S. Towner, T. W. Geisbert, P. B. Jahrling, C. Büchen-Osmond, H. Ellerbrok, M. P. Sanchez-Seco, Y. Lussier, P. Formenty, S. T. Nichol, H. Feldmann, T. Briese, and W. I. Lipkin.** 2007. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg. Infect. Dis.* **13**:73-81.
16. **Palacios, G., Quan, P., Jabado, O. J., Conlan, S., Hirschberg, D. L., Liu, Y., Zhai, J., Renwick, N., Hui, J., Hegyi, H., Grolla, A., Strong, J. E., Towner, J. S., Geisbert, T. W., Jahrling, P. B., Buchen-Osmond, C., Ellerbrok, H., Sanchez-Seco, M. P., Lussier, Y., Formenty, P., Nichol, S. T., Feldmann, H., Briese, T., and Lipkin, W. I.** 2007. Panmicrobial Oligonucleotide Array for Diagnosis of Infectious Diseases. *Emerging Infectious Disease* **Vol. 13**:73-81.
17. **Paliy, O., and R. Agans.** 2012. Application of phylogenetic microarrays to interrogation of human microbiota. *FEMS Microbiol Ecol* **79**:2-11.

18. **Pease, A., D. Solas, S. Fodor, and e. al.** 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA* **91**:5022-6.
19. **Quan, P.-I., G. Palacios, and O. Jabado.** 2007. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. *J Clin Microbiol* **45**:2359-64.
20. **Rajilić-Stojanović, M., H. Heilig, D. Molenaar, K. Kajander, A. Surakka, H. Smidt, and W. de Vos.** 2009. Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**:1736-51.
21. **Schena, M., D. Shalon, P. Brown, and e. al.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**:467-70.
22. **Singh-Gasson, S., R. D. Green, Y. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina.** 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17**:974-978.
23. **Urisman, A., K. Fischer, and C. Chiu.** 2005. E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. *Genome Biology* **6**:R78.
24. **Victoria, J. G., C. Wang, M. S. Jones, C. Jaing, K. McLoughlin, S. Gardner, and E. L. Delwart.** 2010. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J. Virol.* **84**:6033-6040.
25. **Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem, and J. L. DeRisi.** 2002. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* **99**:15687–15692.
26. **Wang, D., A. Urisman, Y.-T. Liu, M. Springer, T. G. Ksiazek, D. D. Erdman, E. R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J. P. Latreille, R. K. Wilson, D. Ganem, and J. L. DeRisi.** 2003. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* **1**(2):e2 doi:10.1371/journal.pbio.0000002.
27. **Wang, J., J. Van Nostrand, L. Wu, Z. He, G. Li, and J. Zhou.** 2011. Microarray-based evaluation of whole-community genome DNA amplification methods. *Appl Environ Microbiol* **77**:4241-5.
28. **Wilson, K. H., Wilson, W. J., Radosevich, J. L., DeSantis, T. Z., Viswanathan, V. S., Kuczmarski, T. A., and Andersen, G. L.** 2002. High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Applied and Environmental Microbiology* **Vol. 68**:0099-2240
29. **Yozwiak, N., P. Skewes-Cox, M. Stenglein, A. Balmaseda, E. Harris, and J. DeRisi.** 2012. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* **6**:e1485.

Table 1. Comparison of whole genome pathogen microarray vs multiplex PCR, phylogenetic microarray and next generation sequencing in terms of resolution, computational time, throughput and cost.

Pathogen detection technology	Resolution	Computational time	Throughput	Cost
<i>Multiplex qPCR</i>	+	+	++++	+
<i>Phylogenetic microarray</i>	++	++	+++	++
<i>Whole genome microarray</i>	+++	++	+++	++
<i>Next-generation sequencing</i>	++++	++++	+	++++

Figure 1. Comparison of microarray, PCR and sequencing in cost, time, resolution and breadth of detection.

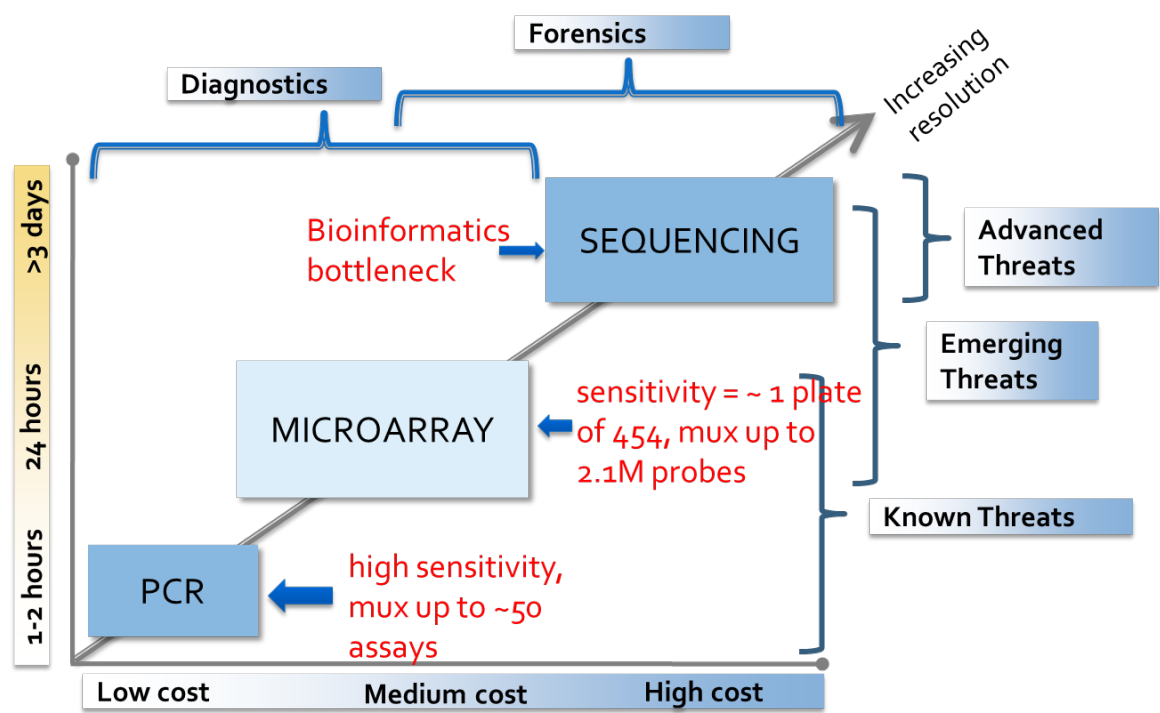


Figure 2. An example of a human clinical sample run on the LLMDA array and analyzed using the Composite Likelihood Maximization method. Two species were detected from this sample, an E coli and a human papillomavirus.

